LA-UR-03-1504
*Approved for public release;
distribution is unlimited.*

Title: **A Theoretical Motivation for Patterns in Information Extraction**

Author(s): Cornelia M. Verspoor, George J. Papcun and Kari Sentz

Submitted to:

http://lib-www.lanl.gov/cgi-bin/getfile?00937082.pdf

A Theoretical Motivation for Patterns in Information Extraction

Cornelia M. Verspoor
verspoor@lanl.gov

George J. Papcun
gjp@lanl.gov

Kari Sentz
ksentz@lanl.gov

Los Alamos National Laboratory
Computer & Computational Science Division
PO Box 1663, MS B256
Los Alamos, NM 87545

Abstract

We argue that the representation of lexical and structural patterns, and specifically patterns incorporating features and constraints from multiple linguistic levels, captures important generalizations about form-meaning correspondence and furthermore suggests fruitful paths for NLP resource development. These arguments are based on the framework of Construction Grammar, and we show how we made use of this framework in developing our own information extraction prototype.

1 Introduction

Approaches to information extraction vary considerably in the depth of linguistic analysis they perform on their inputs. Deep approaches (e.g. Hobbs 1991) attempt full syntactic, semantic, and/or discourse processing as a basis for recognizing relations expressed in texts. Shallower approaches (e.g. Hobbs et al 1996) make use of chunk parsing to identify simple linguistic constituents, defining domain-specific patterns which abstract over those constituents and associating them directly with relational template structures. There are solutions that exist in between these two extremes, such as the LaSIE system (Humphreys et al. 1998) which makes use of some finite-state pattern recognition but also incorporates parsing coupled with compositional semantic interpretation rules and discourse modeling, or the WHITEBOARD system (Crysmann et al, 2002) which integrates shallow analyses with deep parsing by pre-compiling representations for certain sequences, and by mapping structures identified in shallow processing to HPSG types.

The arguments made for the strategy in designing an information extraction system normally center on performance issues. Shallower approaches are more robust to the linguistic variance (both syntactic and lexical) of free text, they are much faster (both in throughput and development time), and they have been proven to achieve good results on information extraction tasks. Deeper approaches support richer semantics and a more precise analysis, are capable of supporting recognition of more complex relational structures, and are in principle more domain-neutral because they embody general linguistic principles. Few attempts have been made to motivate shallow approaches in linguistic terms. The attitude in the IE community appears to be that creating gazetteers (lists of names) to help in named entity recognition or defining patterns which represent abstractions of how an entity or relation is expressed in the domain of interest is somehow not linguistics, but that it represents a valuable language engineering approach to the problem in spite of lacking a theoretical basis.

We argue in this paper that in fact there is a theoretical motivation for the shallow approach represented by the vast majority of implemented information extraction systems, in the form of Construction Grammar (cf. Fillmore, 1985), and we will describe a prototype system that we have been building to extract protein/gene interactions by following the precepts of Construction Grammar. We will further discuss why it is important for research in information extraction that such a theoretical motivation exists.

2 Construction Grammar

Construction Grammar (Fillmore, 1985; Fillmore et al, 1988; Goldberg, 1995) is a framework for linguistic analysis in which it is argued that language consists of a set of patterns at varying levels of abstraction that integrate form and meaning in conventionalized and often non-compositional ways.

We define a construction as any learned relationship between form and meaning in a language. Following (Goldberg, 1995), we state this formally as follows (see also Papcun et al, 2003):

`C is a construction iff C is a form-meaning pair <` F_i`,` S_j`>, such that some aspect of` F_i `(form) or some aspect of` S_j `(semantics) is not strictly predicted from C's component parts or from other previously established constructions.`

This definition includes a spectrum of elements ranging from the most specific to abstract and schematic grammatical constructions. A grammar for a language consists of a collection of such patterns, which provide a direct mapping from forms to meanings and vice versa.

- words: `table`, `chair`
- morphemes: `red+ness, swift+ly`
- word compounds: `ranch house`, `dog house`, `Frank Lloyd Wright house`
- idioms: `kick the bucket, let the cat out of the bag`
- entrenched collocations with fixed meanings: X `varies as a function of` Y `it is shown that`
- schematic grammatical constructions

The CG approach was developed in response to the need to account for constructions such as open idioms that cannot be adequately accounted for in other grammar theories, and to represent such marked constructions in the same formal system as "regular" patterns of the language. It has been bolstered by evidence from child language acquisition (cf. Tomasello, 1998) that suggests that children accumulate phrases that are initially unanalyzed, and never acquire a compositional analysis for many phrases.

2.1 Constructions as primary linguistic units

Following Construction Grammar, we therefore claim that constructions can be considered primary linguistic units, and that it is not necessary, and often undesirable, to attempt a deep analysis of some phrases or sentences. An attempt to devise a grammar or construct a system which can provide a deep analysis of all sentences in fact runs into several difficulties beyond issues of robustness and throughput:

1. Certain constructions violate 'expected' argument structures, for instance introducing an additional complement and some additional component of meaning which is not normally attributed to the head word of the phrase. As an example, in a caused motion construction (as in "`John sneezed the tissue off the table`") a normally intransitive verb requires a direct object and a PP complement and acquires the caused-motion interpretation. Goldberg (1995) has argued persuasively against postulating additional verb senses (with alternative subcategorization frames) to explain such phenomena.
2. Many conventionalized phrases, such as those identified in Named Entity recognition (e.g. "`The New York Times`"), have many possible syntactic analyses, only one of which makes sense given the known meaning. Generating all of the possible analyses during parsing is computationally expensive and highly error-prone unless guided by pragmatic constraints or world knowledge.

There are simply many phrases whose meaning or structure cannot be adequately explained on the basis of their constituent structure. Often, it is not worth attempting to perform deep analysis on them, as that analysis does not necessarily lead to the interpretation or implications of that phrase. Consider a product name such as "I Can't Believe It's Not Butter" – what does one gain by performing a full syntactico-semantic analysis of that phrase (cf. Appelt and Israel, 1999)? Computationally, it is impractical, but more importantly, the internal structure of such phrases is not relevant to their final interpretation. There is thus a linguistic motivation for representing such phrases as units – they have a consistent interpretation which is not predictable from their constituents.

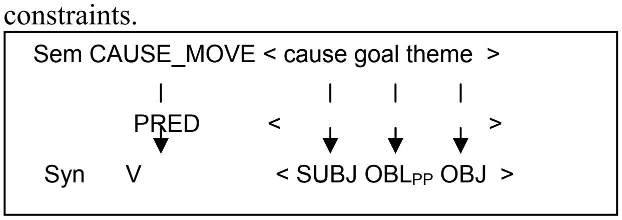
2.2 Compositionality

It is important to note that Construction Grammar does not reject the notion of compositionality in its entirety. Instead, it adopts a weaker form than is traditionally assumed in formal semantics by stating that the meaning of an expression results from integration of the meanings of lexical items or other constituents with the meanings of constructions as a whole (Goldberg, 1995). This weaker notion of compositionality resolves the apparent tension between the Construction Grammar view of constructions as corresponding to basic units of meaning in language, and the explanatory power of compositionality in much analysis – the attribution of meaning to a construction as a whole does not deny the meaning contribution of the constituents of those constructions. Rather, Construction Grammar leads us to recognize that the meaning of a whole may only be partially predicted from the meanings of the parts, and provides us with a formal mechanism for modeling when this is the case, while also allowing for the meaning of a whole to have little to do with the meanings of its parts (as for idioms which may have an available compositional diachronic interpretation but no obvious compositional synchronic explanation), or only to do with the meaning of its parts (as in simple sentences in which a strictly lexical semantic account of the interpretation is the most straightforward explanation).

By acknowledging the status of the construction as a linguistic unit with its own meaning and its own constraints on usage, Construction Grammar gives us a framework which relaxes the need to drive linguistic analysis entirely from the lexicon, and allows for the possibility of pragmatic constraints on interpretation (cf. Verspoor, 1997). It is therefore a more integrative approach to modeling language use and interpretation than theories that expect a more strictly compositional analysis.

2.3 An abstract construction

As an example of what we mean by a schematic or abstract construction, let us consider Goldberg's (1995) representation of the caused motion construction. This representation couples a syntactic frame with a semantic frame,[1] but one can also imagine constructions which add pragmatic usage constraints.



The construction indicates that the lexical semantic frame of a verbal predicate is integrated with semantic frame of the construction, such that the subject acquires a cause role, the oblique prepositional phrase a goal role, and the object a theme role. In this case, the verbal predicate will not usually subcategorize for the oblique and direct object arguments, so the semantics of those constituents will come solely from the construction itself, and it is the construction which licenses those arguments. If the verbal predicate subcategorizes for the subject, then the cause role will be attributed to that subject in addition to whatever semantics is attributed by the predicate itself.

In this example, one can see that constructions can consist of syntactic labels and make use of lexical semantic structures. Thus it is clear that CG in itself does not argue against "deeper" linguistic processing or representation of linguistic information at the lexical or syntactic levels. However, it provides motivation for attributing meaning directly to larger structures, which may have "deeper" constituents, or which may be "shallow" and consist of fixed phrases or some combination of constraints at these different levels of abstraction. Since it allows for both, it does not assume "shallow" analysis to the exclusion of "deep" analysis, nor vice versa.

3 Designing an IE system based on CG

Our task was to develop a prototype system for extracting gene/protein interactions, involving a predefined set of genes and proteins of interest, from biological texts. The approach we took to design the prototype is motivated by the Construction Grammar framework, yet is very similar to the approach taken in most IE systems. The architecture is shown in Figure 1. See also Papcun et al (2003).

Since our gazetteers for genes and proteins (referred to as *factors* by our domain experts) were defined in advance, and we were not concerned with distinguishing ambiguous cases, our main focus was on linguistic analysis of relations of interest. We were given a corpus of approximately 600 sentences for which desired target outputs in the form of Agent-Action-Patient relations (AAPs) had been defined by a domain expert. This was used as a basis for identifying all the words which expressed the actions of interest.

We created two lists relevant to relation identification, in addition to the factor list of gene and protein names. One contains verbal forms of the verbs expressing actions of interest, the second contains

[1] Researchers in Construction Grammar do not currently make use of a standard formalism, although Kay (2002) has adopted Minimal Recursion Semantics (Copestake et al. 1995).

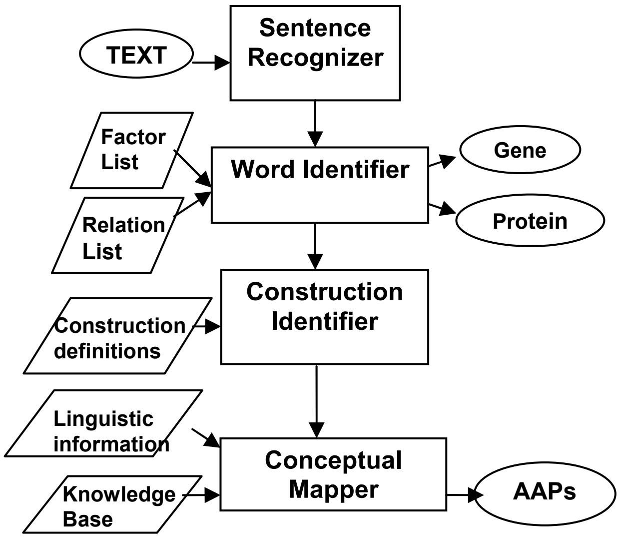


Figure 1: Architecture of the prototype gene/protein interaction extraction system.

nominalizations of those verbs. Samples of these lists are shown below. The lists reflect domain-specific constructions at the word level.

FACTORS	ACTIONS	NOMINALS
tttf1	activate	activation
fos	inhibit	inhibition
camk1	regulate	regulation
camkk	cotransfect	cotransfection
diacylglycerol	phosphorylate	phosphorylation
prkc	stimulate	stimulation

The manual analysis of the corpus also resulted in the identification of frequently occurring constructions through which the relations were expressed in the corpus. We focused on three main constructions for the initial prototype: active, passive, and nominalization constructions.

3.1 Active construction

The active construction is defined as a construction in which the subject performs the action indicated by the verb. The sentence "Our results define a cycle in which diacylglycerol activates protein kinase C, which then regulates the metabolism of diacylglycerol by alternating the intracellular location of dgkz." (Topham et al, 1998) matches:

diacylglycerol	activates	protein kinase C
Factor	**Active Verb**	**Factor**

This construction maps to this meaning frame:

diacylglycerol	activates	protein kinase C
Agent Factor	**Action**	**Patient Factor**

This construction corresponds to the transitive verb frame and is highly useful in our task. This construction serves the same essential function as a transitive subcategorization frame with associated semantic frame at the lexical level, but we have generalized it to allow for unrecognized intervening words and phrases. The advantage of using a construction here is that captures the general constraint that arguments must be factors in order to map to the associated meaning frame. Furthermore, our implementation allows us to pick out this pattern in a complex sentence without having a full grammar capable of handling intervening modifiers.

3.2 Passive construction

Passive constructions are those in which a patient is expressed as the subject and an agent is expressed as the object of the preposition "by". So "Several recent studies have shown that camk1 (camk1) is phosphorylated and activated by a protein kinase (camkk) that is itself subject to regulation by ca2+/calm." (Matsushita, Nairn, 1999) matches:

camk1	is activated	by	camkk
Factor	**Passive Verb Group**	**"by" prep**	**Factor**

This construction maps to this meaning frame:

camkk	activates	camk1
Agent Factor	**Action**	**Patient Factor**

In a lexical semantic account, the passive construction is normally handled via a rule which generates the alternate passive subcategorization frame for transitive verbs. However, this approach assumes that all the verbs of interest exist in the lexicon with transitive frames to which the lexical rule can be applied. Using a construction, in contrast, we are able to recognize the pattern and understand the semantic roles the constituents of the pattern play, independently of knowledge of the verb. Perhaps more importantly, a lexical-rule based account does not adequately capture the pragmatic conditions that are associated with the use of a passive, while a constructional account allows us to directly associate those conditions with the pattern since the pattern is explicitly represented as a linguistic unit with its own linguistic properties (cf. Croft, 1991).

3.3 Nominalization construction

Nominalizations are constructions that use a noun that is related to a verb. For “These results suggest that a pik3/p70(s6k)-dependent pathway is required for regulation of HKII gene transcription by ins and that the kras2-mapk-dependent pathway is probably not involved.” (Osawa et al, 1996) we match:

regulation	of	HKII	by	ins
Nominal	**“of” prep**	**Factor**	**“by” prep**	**Factor**

This construction maps to this meaning frame:

ins	regulates	HKII
Agent Factor	**Action**	**Patient Factor**

Again, the formal representation of this pattern allows us to directly attribute meaning to the construction, independently of the individual words which instantiate the pattern. We do not have to assume a complex subcategorization frame for the nominals, which is repeated for each individual nominal and fails to capture the overall generalization about the interpretation of the pattern.[2]

In the case of our nominalization construction, by restricting the arguments of the prepositions to be factors, we can associate a specific causal interpretation with the construction which maps into the active AAP relational form shown above. That is, any occurrence of the pattern “*ion of <$factor_1$> by <$factor_2$>” can be interpreted as “$factor_2$ <verb>s $factor_1$”, or “$factor_2$ causes $factor_1$ to be <verb>ed”. Note that this interpretation does not extend to many superficially similar occurrences of this pattern, such as “`the administration of medicine by IV`” which does not mean “`(the) IV administered the medicine`” in the most salient interpretation, but rather is interpreted as indicating that the IV was the means of administration of the medicine. Contrast this with “`the administration of medicine by the nurse`” which does have the relevant causal interpretation. By placing appropriate semantic constraints on the constituents of the constructions, such meaning variations can be captured in a general and productive way without assuming proliferation of lexical entries.

[2] See Zadrozny (1995) for arguments about the compactness of construction-based representations as compared to lexicalized representations.

3.4 Processing

During processing, a text is segmented into individual sentences and each sentence is analyzed for the presence of elements from the factor and relation lists. Any elements found will be marked up and the sentence will be passed to the Construction Identifier component which attempts to match the structure of the sentence to one of the known constructional patterns. If a pattern is recognized by the Construction Identifier component, the constituents of the construction are extracted and the Conceptual Mapper generates normalized AAP relations. This mapping is performed using linguistic knowledge of the correspondence of constructions to the formal representation of the meaning expressed by the construction, as well as lexical knowledge about the relationships between verb forms and nominal forms of actions. In future work, we plan to make use of more sophisticated domain knowledge to improve precision at the mapping stage by verifying additional semantic constraints on individual constructions instantiated by specific lexical items.

4 Implications for IE and beyond

4.1 A motivation for existing solutions

As suggested previously, the use of gazetteers in information extraction is motivated by a linguistic argument about conventionality in language. By similar reasoning, there is little to gain by analyzing structures such as dates, numbers, monetary expressions, etc. and there is linguistic support for the definition of recognizers for such consistently structured expressions because there is no principled way to capture their overall interpretation compositionally.

It has been argued that gazetteers and specific phrase recognizers are of limited utility in the face of free text, where there will always exist names not listed in the gazetteer or patterns not conforming to the constructions defined in recognizers, despite the fact that they capture conventionalized language use in particular domains. The solutions to this problem, in the form of heuristic rules based on context (e.g. Mikheev et al, 1998), or machine learning models which generalize from examples (e.g. Baluja et al, 1999), further emphasize the constructional nature of language, since it is possible to characterize the sentential or discourse contexts in which specific types of words or phrases occur with high accuracy.

The effectiveness of techniques which make use of lexical resources or phrasal/sentential patterns extends beyond information extraction. The phrasal lexicon approach to natural language generation (Milosavljevic et al, 1996) makes use of complex elements in their knowledge representation that are mapped directly to linguistic elements which are correspondingly complex. For instance, the lexical item "is a carnivore and eats ants, termites and earthworms" is postulated to correspond to the concept `eats-ants-termites-earthworms`, rather than requiring the construction of the phrase from its consituents. This approach is motivated in terms of reuse and efficiency. They state, "If we repeatedly realise a semantic element in the same way, it is better to remember this and avoid rebuilding the surface form each time". We suggest that the construction of a phrasal lexicon also observes the deeper principle of conventionality in language use.

Zadrozny et al (1994) present a natural language speech-enabled interface to a calendar application that has an architecture motivated explicitly in terms of constructions, and utilize a parser which is able to take into account form, meaning, and context simultaneously by having these different levels integrated into a common representation, and using this representation for not only lexical items, but also for syntactic forms and sentential patterns. They argue that this architecture allows them to build a system which is discourse-oriented, taking context into account during interpretation, and show that it is feasible to build a working system in terms of constructions.

Construction Grammar motivates the development of NLP systems which capture the conventionality of language and avoid attributing unnecessary ambiguity to conventionalized linguistic structures. Specifically,

- Gazetteers and other lexical resources capture domain-relevant words or phrases which should be treated as a unit, syntactically and semantically.
- The definition of domain patterns is not simply an efficient means to an ends for representing common ways in which information is expressed. These patterns capture manners of expression in particular domains which are consistent, powerful models of linguistic competence in the domain and can very often be argued to carry meaning or implications which are not strictly apparent from compositional analysis of the constituents. Furthermore, domain patterns enable us to characterize fine-grained semantic and pragmatic conditions which license their applicability and enable us to avoid extreme lexical redundancy (Zadrozny, 1995).
- The development of heuristics or models based on properties of the linguistic context, such as for semantic discrimination tasks like named entity recognition, reflects the CG-justified property of multiple linguistic levels interacting in establishing meaning. Characterization of a phrase as a name or location through context rules which refer to words as well as syntactic and semantic labels (e.g. the rule in the Mikheev et al 1998 work which establishes X as a person name via the pattern [X is a {JJ}* PROFESSION]) reflects this important linguistic property.

4.2 A framework for future solutions

By recognizing the theoretical motivations for an approach to information extraction (and indeed any NLP system which hopes to model human competence in language understanding to one degree or another) in which conventionalized phrases are repesented and used without deeper analysis, we can gain a framework for guiding the further development of innovative solutions. We suggest that the construction grammar framework provides important insights into resource creation for NLP systems. In particular, Construction Grammar motivates at least the following NLP tasks:

1. automatic gazetteer development
2. automatic mining of recurrent patterns spanning linguistic levels

Waterman (1993) argued that little progress has been made in automating the acquisition of lexicons for pattern-based IE systems due to the lack of a linguistic framework in which to view their use. He continues by suggesting that a lexical semantic framework can provide the needed linguistic grounding, and shows how clustering of phrases containing particular common words can be effective for identifying domain collocations. However, while we agree with Waterman's contention that a linguistic framework is required to drive automatic acquisition of patterns for IE, his outlined approach is not a sufficient solution as it offers no mechanism for abstracting phrase clusters into patterns, or for incorporating syntax or semantics into those patterns. The arguments in favor of Construction Grammar make clear that a more holistic approach to representing linguistic structure is needed.

Recently, there has been a growing interest in automatic techniques for inducing syntactico-semantic patterns from text. Initial efforts focused on syntax, specifically subcategorization frames (Manning, 1993; Briscoe and Carroll, 1997) and grammar induction (Klein and Manning, 2002), but these approaches are being extended to incorporate semantic features (Grishman and Sterling, 1992; Zechner, 1998).

What Construction Grammar tells us is that we can and must go further in mining domain patterns, because these patterns form an important part of linguistic competency. Such patterns must be recognized as linguistic units which pervade language use. Specifically, we suggest that a research path which focuses on trying to identify recurrent constructions in a domain corpus, with increasingly sophisticated linguistic generalizations and linguistic constraints (lexical, syntactic, semantic, or pragmatic), is likely to be particularly fruitful for development of powerful, retargetable IE applications. Such a research path would naturally start with mining collocations from corpora (cf. Smadja, 1993), and would focus on developing techniques for increasing the abstractions represented in those collocations – incorporating gaps (probably constrained), part of speech tags, syntactic chunks, semantic tags, etc. such that the patterns ultimately capture the constructions of the language.

Finally, CG provides us with a linguistic framework with which to evaluate domain patterns. In the context of CG, we must treat domain patterns as linguistic entities, and formally represent the meaning they contribute and the linguistic constraints they incorporate. We can therefore assess the extent to which specific patterns have special meanings or fine-grained constraints, and contrast that assessment with what would be predicted through strictly compositional interpretation. To the degree that domain patterns have specific linguistic properties that are unique to the patterns, they can be validated as independent linguistic units.

5 Conclusion

We have provided a new perspective on the use of gazetteers, domain patterns, and context rules in information extraction. We have argued that in addition to being convenient resources for language engineering, such resources are explicitly motivated in theoretical terms. Their theoretical grounding comes from the Construction Grammar framework.

We have shown how domain patterns can be viewed as constructions through the example of our prototype biological information extraction system, and how the constraints on those patterns are important for establishing their meaning.

The Construction Grammar framework can be drawn on to motivate “shallow” approaches to information extraction, but as a result of this motivation it could be argued that these approaches should no longer be given the “shallow” attribution. Instead, we suggest that these techniques more naturally and accurately reflect linguistic competence than “deep” approaches which are lexically driven, compositional, and require sophisticated grammars. This is not to say that we reject the contributions of lexical semantics, compositionality, or grammar. Rather, we suggest that they can most effectively be utilized in conjunction with constructional representations. Adopting this perspective provides a powerful framework for driving research in NLP to use representations integrating features and constraints at multiple levels of linguistic abstraction.

6 Acknowledgements

This work was funded in part through a Los Alamos National Laboratory collaboration with Procter & Gamble. We would like thank P&G for providing us with data and their domain expertise.

References

Appelt, D. and D.J. Israel. 1999. ``Introduction to Information Extraction Technology''. Tutorial at *IJCAI-99*.

Baluja, S., V. Mittal, and R. Sukthankar, "Applying Machine Learning for High Performance Named-Entity Extraction," *Pacific Association for Computational Linguistics*, 1999.

Briscoe, T. and J. Carroll. 1997. Automatic extraction of subcategorization from corpora. In *Proceedings of the 5th ANLP Conference*, Washington, DC, pages 24-29.

Copestake, A., D. Flickenger, R. Malouf, S. Riehemann, I. Sag. 1995. Translation Using Minimal Recursion Semantics. In Proceedings of the 6th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-95). Leuven, Belgium.

Croft, W. 1991. *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. The University of Chicago Press, Chicago, IL.

Crysmann, B., A. Frank, B. Kiefer, S. Muller, G. Neumann, J. Pskorski, U. Schafer, M. Siegel, H. Uszkoreit, F. Xu, M. Becker, H. Krieger. 2002. An integrated architecture for shallow and deep processing. *Proceedings of the 40th Annual Meeting of the ACL*, pages 441-448.

Fillmore, C. 1985. Syntactic intrusion and the notion of grammatical construction. *Berkeley Linguistics Society* 11: 73-86.

Fillmore, C., P. Kay, and M.C. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions. *Language*, 64(3):501-538, 1988.

Grishman, R., and Sterling, J. 1992. Acquisition of selectional patterns. In *COLING'92, the Proceedings of the 14th International Conf. on Computational Linguistics*, Nantes, France.

Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Hobbs, J.R., D.E. Appelt, J. Bear, D. Israel, M. Kameyama, M. Stickel, and M. Tyson. 1996. FASTUS: A cascaded finite-state transducer for extracting information from natural-language text. In *Finite State Devices for Natural Language Processing*. MIT Press, Cambridge, MA.

Hobbs, Jerry R. 1991. Description of the TACITUS system as used for MUC-3. In *Proceedings of the Third Message Understanding Conference MUC-3*, pages 200-206. Morgan Kaufman.

Humphreys, K., R. Gaizauskas, S. Azzam, C. Huyck, B. Mitchell, H. Cunningham, Y. Wilks. 1998. University of Sheffield: Description of the LaSIE-II System as Used for MUC-7. In *Proceedings of the Seventh Message Understanding Conference (MUC-7).*

Kay, P. 2002. An Informal Sketch of a Formal Architecture for Construction Grammar. *Grammars* 5:1.

Klein, D. and C. Manning. 2002. A Generative Constituent-Context Model for Improved Grammar Induction. In *Proceedings of the 40th Annual Meeting of the ACL*, pages 128-135.

Manning, C. 1993. Automatic acquisition of a large subcategorization dictionary from corpora. In *Proceedings of the 31st Annual Meeting of the ACL*, pages 235-242.

Matsushita M, Nairn AC. (1999) "Inhibition of the Ca2+/calmodulin-dependent protein kinase I cascade by cAMP-dependent protein kinase." *J Biol Chem* Apr 9, 1999; 274(15):10086-93.

Osawa H, Sutherland C, Robey RB, Printz RL, Granner DK. 1996. "Analysis of the signaling pathway involved in the regulation of hexokinase II gene transcription by insulin." *J Biol Chem*; 271(28):16690-4

Milosavljevic, M., A. Tulloch and R. Dale. 1996. Text Generation in a Dynamic Hypertext Environment. Pages 417-426 in *Proceedings of the 19th Australasian Computer Science Conference*, Melbourne, Australia, 31 January-2 February.

Mikheev, A., C. Grover and M. Moens (1998): Description of the LTG System Used for MUC-7. In *Proceedings of 7th Message Understanding Conference*.

Papcun, George, Kari Sentz, Andy Fulmer, Jun Xu, Olaf Lubeck, and Murray Wolinsky. 2003. A construction grammar approach to extracting regulatory relationships from biological literature. *Pacific Symposium on Biocomputing 2003*, Kauai, Hawaii.

Smadja, F. 1993. Retrieving collocations from text: Xtract. In *Computational Linguistics* 19:1.

Tomasello, M., Ed. 1998. *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure.* Erlbaum.

Topham MK, Bunting M, Zimmerman GA, McIntyre TM, Blackshear PJ, Prescott SM. 1998. Protein kinase C regulates the nuclear localization of diacylglycerol kinase-zeta. *Nature*; 394(6694):697-700.

Verspoor, C. 1997. *Contextually-Dependent Lexical Semantics*. PhD thesis, the University of Edinburgh.

Waterman, S. 1993. Structural Methods for Lexical/Semantic Patterns. In B. Boguraev and J. Pustejovsky, editors, *Proceedings SIGLEX Workshop: Aquisition of Lexical Knowledge from Text*, pp. 128-142. Conference of the ACL, Ohio State University.

Zadrozny, W., M. Szummer, S. Jarecki, D. Johnson, L. Morgenstern. 1994. NL Understanding with a Grammar of Constructions. *15th Intl. Conf. on Computational Linguistics (COLING)*, pages 1289-1293.

Zadrozny, W. 1995. The Compactness of Construction Grammars. IBM Research Report RC 20003 (88493).

Zechner, K. 1998. Automatic construction of frame representations for spontaneous speech in unrestricted domains. In *Proceedings of COLING/ACL 98*.